# ECE 411 Introduction to Machine Learning
# Fall 2023 Exam 2
# Instructor: Dr. Chau-Wai Wong

This is a closed-book exam. You may use a scientific calculator with cleared memory, but not a smart phone or computer. Two-sided letter-sized handwritten cheatsheet is allowed. You should answer ***all four*** problems.

**Problem 1** (25 pts) This problem concerns the variance operator.

**(a)** You are given random variables $\{X_i\}_{i=1}^n$ and nonzero constants $\{a_i\}_{i=1}^n$. Prove the following equality:

$$\mathrm{Var}(a_1X_1 + a_2X_2) = a_1^2\mathrm{Var}(X_1) + a_2^2\mathrm{Var}(X_2) + 2a_1a_2\,\mathrm{Cov}(X_1, X_2). \tag{1}$$

There are multiple proving routes. You may find some of the following expressions useful: i) the definitions

$$\mathrm{Var}(X) = \mathbb{E}[(X - \mathbb{E}X)^2], \tag{2a}$$

$$\mathrm{Cov}(X, Y) = \mathbb{E}[(X - \mathbb{E}X)(Y - \mathbb{E}Y)], \tag{2b}$$

ii) the shortcut formulas

$$\mathrm{Var}(X) = \mathbb{E}[X^2] - (\mathbb{E}X)^2, \tag{3a}$$

$$\mathrm{Cov}(X, Y) = \mathbb{E}[XY] - \mathbb{E}[X]\,\mathbb{E}[Y]. \tag{3b}$$

**(b)** An ECE student named Tom plans to test the fuel economy of his car in terms of how many gallons is needed for driving one mile. He will do four test drives of $x_i$ miles each, $i = 1, \ldots, 4$, and will measure the corresponding gas consumption $Y_i$ gallons, $i = 1, \ldots, 4$ using a meter connected to his car's microcontroller, where $Y_i$ has a variance of $\sigma^2$ for all $i$. Denote the ground-truth fuel economy as $k$ gallon/mile. Tom and his friends used intuition and applied some principled approaches and came up with the following candidate estimators for the fuel economy $k$:

$$\hat{k} = \left(\sum_{i=1}^{4} x_i Y_i\right) \Bigg/ \left(\sum_{i=1}^{4} x_i^2\right), \tag{4a}$$

$$\tilde{k} = \frac{1}{4}\sum_{i=1}^{4} Y_i/x_i, \tag{4b}$$

$$\check{k} = \left(\sum_{i=1}^{4} Y_i\right) \Bigg/ \left(\sum_{i=1}^{4} x_i\right). \tag{4c}$$

(i) Derive the analytic forms of the variance of estimators $\hat{k}$, $\tilde{k}$, and $\check{k}$.

(ii) Let $(x_1, x_2, x_3, x_4) = (2, 1, 2, 3)$. Numerically calculate the variance of all estimators.

(iii) In Exam 1, you have proved that all three estimators are unbiased estimators. Based on the results in (ii), argue which one may be the best for Tom and his friends to use. Please give justification in your own words.

**Problem 2** (25 pts) This problem concerns the maximum likelihood estimator (MLE).

**(a)** Calculate the MLE for variance $\sigma^2$ for a random sample $X_1, \ldots, X_n$ drawn from the normal distribution with the PDF shown as follows:

$$f(x; \sigma^2) = \frac{1}{\sqrt{2\pi}\sigma} e^{-(x-\mu)^2/2\sigma^2}, \quad -\infty < x < \infty. \tag{5}$$

**(b)** Calculate the MLE for parameter $b$ for a random sample $X_1, \ldots, X_n$ drawn from an exponential distribution with the PDF of the following form:

$$f(x; \lambda) = \lambda e^{-\lambda x}, \quad x \geq 0. \tag{6}$$

You may find the invariance principle of MLE handy.

**Problem 3** (25 pts) This problem concerns the curse of dimensionality.

**(a)** Suppose that we have a set of observations, each with measurements on $p = 1$ feature, $X$. We assume that $X$ is uniformly distributed on $[-1, 1]$. Associated with each observation is a response value. Suppose that we wish to predict a test observation's response using only observations that are within 10% of the range of $X$ closest to that test observation. For instance, in order to predict the response for a test observation with $X = 0.2$, we will use observations in the range $[0.1, 0.3]$. On average, what fraction of the available observations will we use to make the prediction for a test observation with $X = 0.3$?

**(b)** Now suppose that we have a set of observations, each with measurements on $p = 2$ features, $X_1$ and $X_2$. We assume that $(X_1, X_2)$ are uniformly distributed on $[-1, 1] \times [-1, 1]$. We wish to predict a test observation's response using only observations that are within 20% of the range of $X_1$ and within 20% of the range of $X_2$ closest to that test observation. On average, what fraction of the available observations will we use to make the prediction for a test observation with $(X_1, X_2) = (0.1\pi, 0.2\pi)$?

**(c)** Generalize the cases in (a) and (b) to $p = 100$. What fraction of the available observations will we use to make the prediction? We assume that the test observation stays far away from the boundaries.

**(d)** Using your answers to (a)–(c), comment on a drawback of $k$-NN when $p$ is large.

**(e)** Now suppose that we wish to make a prediction for a test observation by creating a $p$-dimensional hypercube centered around the test observation that contains, on average, 20% of the training observations. For $p = 1, 2,$ and 100, what is the length of each side of the hypercube?

**Problem 4** (25 pts) This problem concerns the application of the method of nearest neighbors. A set of 6 training data points is drawn in the given figure. Using the $k$-nearest-neighbor regression rule, an estimated regression function can be written as follows:

$$\hat{y}^{(k)} = \hat{f}^{(k)}(x) = \frac{1}{k} \sum_{i \in I(x)} y_i, \quad I(x) = \{i : \text{the indices of } k \text{ smallest } |x_i - x|\}. \tag{7}$$

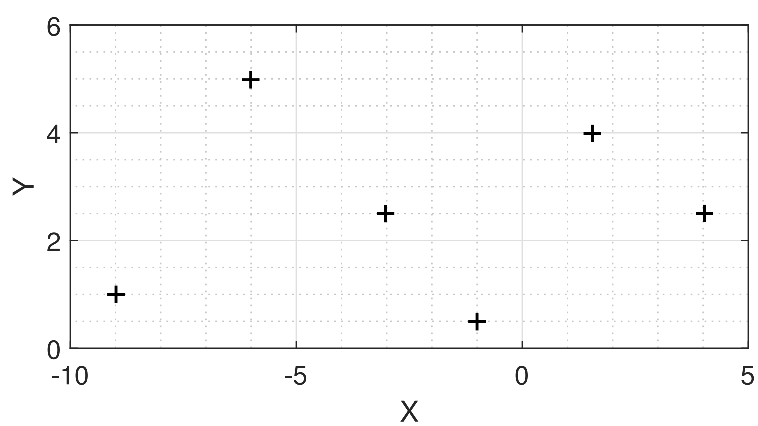
Note that this is a regression problem with only one predictor and the graphical representation of the estimated regression function on the $xy$-plane will be a collection of horizontal line segments.

**(a)** Draw the regression function $\hat{f}^{(k)}(x)$ for $x \in [-10, 5]$ when only $k = 1$ nearest neighbor is contributing to the regression.

**(b)** Draw the regression function $\hat{f}^{(k)}(x)$ for $x \in [-10, 5]$ when two $k = 2$ nearest neighbors are contributing to the regression.

**(c)** Comment on how the shape of regression function will change as the number of contributing neighbors increases.

To get full points, you must annotate the locations of the ***discontinuities*** of each estimated regression function using vertical dotted lines.



[This Page Intentionally Left Blank]

Answer sheet for Problem 1 Name: Obtained Points:

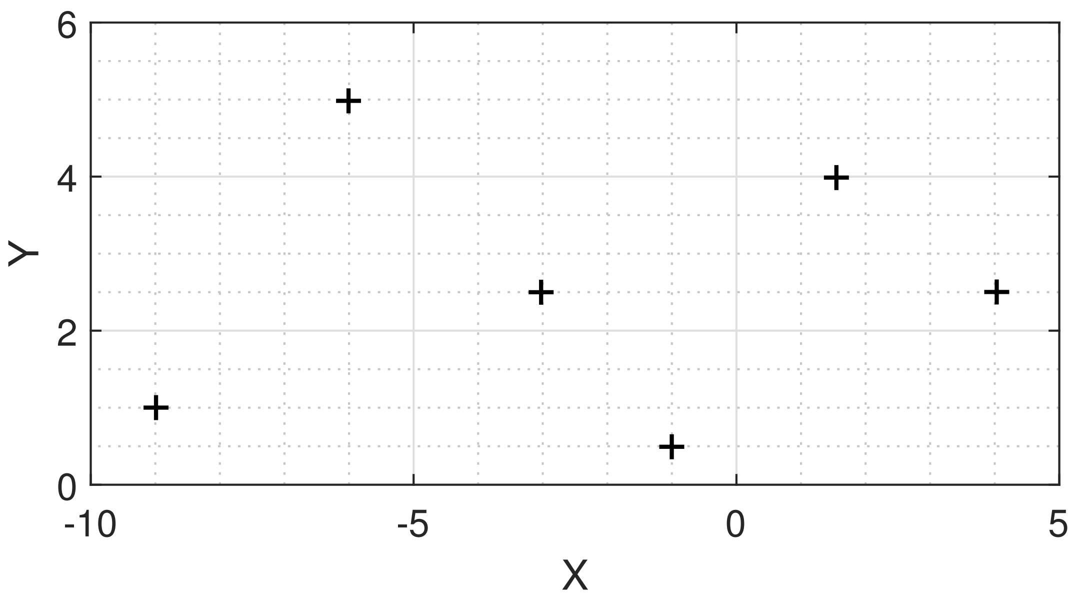
Answer sheet for Problem 1

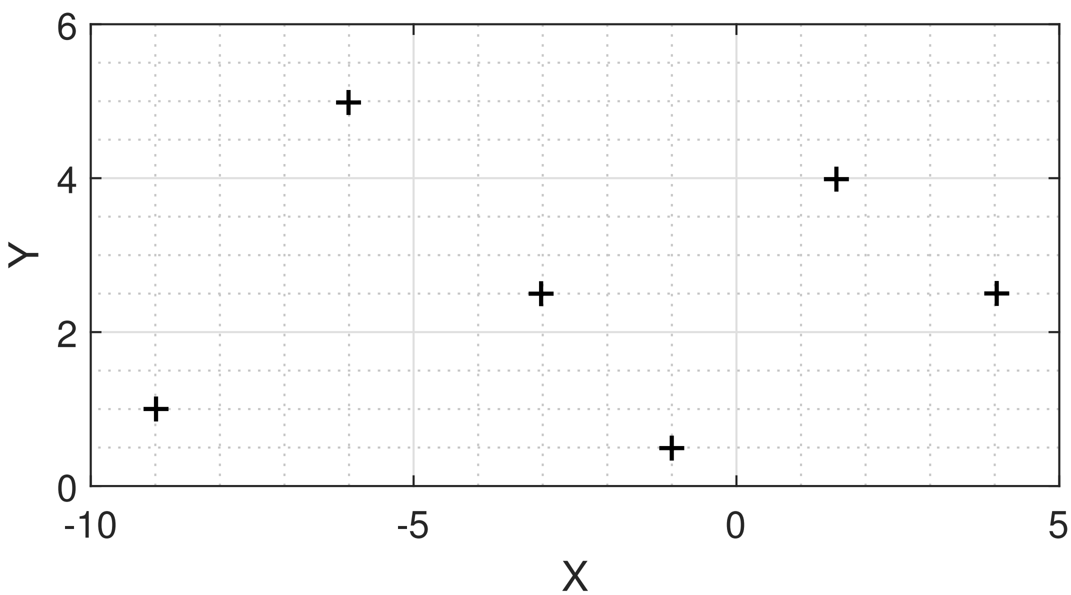
Answer sheet for Problem 2 Name: Obtained Points:

Answer sheet for Problem 2

Answer sheet for Problem 3 Name: Obtained Points:

Answer sheet for Problem 3


Y
X
6
4
2
0
-10
-5
0
5


Y
X
6
4
2
0
-10
-5
0
5

Answer sheet for Problem 4